ORTHORECTIFICATION AND DSM GENERATION WITH ALOS-PRISM DATA IN URBAN AREAS

Thomas Krauß, Mathias Schneider, Peter Reinartz

German Aerospace Center, DLR
Remote Sensing Technology Institute
Münchner Str. 20, D-82234 Weßling
thomas.krauss@dlr.de

ABSTRACT

In this paper different methods for deriving digital surface models (DSM) from ALOS Prism three line stereo images are generated and analyzed. The methods used are classical hierarchical stereo matching with forward intersection and two different dense stereo methods. These are digital line warping which was derived from speech recognition algorithms and semi global matching which is originating in computer vision. All these dense stereo methods need epipolar imagery as input and provide so called disparity images as output. For this in a first step the Prism images has to be transformed by pairs to epipolar geometry. For the reprojection of the disparity images to real DSMs rational polynomial coefficients – which were computed from the satellite ephemeris and attitude date – are used. Finally the DSMs generated by all these different methods are compared to a DSM derived from an Ikonos stereo image pair with a ground sampling distance of 1 m.

1 INTRODUCTION

The Japanese satellite ALOS carries the Prism three line scanner instrument which provides high resolution panchromatic optical images. These images of about 2.5 m ground sampling distance are acquired in the same orbit in nadir and respectively 23.8° forward and backward directions. Based on these images high resolution digital elevation models (DEM) resembling the visible surface (better called digital surface models, DSM) in ground sampling distances of about 2.5 to 10 m can be derived.

Due to the three-ray geometry of the sensor these DSMs can be calculated independently for three image pairs (forward-nadir, nadir-backward and forward-backward) which allows a better elimination of outliers produced by the DSM generation algorithms beyond a simple statistical approach.

For the generation of DSMs exist already a broad repertory of algorithms. Mostly used in photogrammetric standard software is the hierarchical area based matching as described [1]. Recently also algorithms stemming from computer vision applications get more and more common in photogrammetry. First and meanwhile also integrated to standard software was the graph cut [2]. This graph based maximum flow/minimum cut algorithm allows the processing of many input images simultaneously. But due to the bad performance with large images [3] it was not included in this investigation.

A selection of actual dense stereo algorithms used in computer vision can be found in [4]. In our analysis besides the classical method two dense stereo methods get analyzed: First the so called semi global matching introduced by [5]. Second a fused algorithm based on semi global matching and digital line warping. This in turn is based on speech recognition algorithms as known for a long time but only recently adapted to image correlation [6].

The enhancement of the generated DSM is based on detection of so called occlusions in the stereo matching process [7]. An occlusion is an area in the generated DSM which can not be derived from the provided stereo images. In general areas which are not visible simultaneously in both stereo images like small streets or bottoms of steep walls can not be matched and result in so called occluded areas.

Such occlusions can be detected in the DSM generation process of dense stereo algorithms. To fill such occlusions in most cases the assumption that these emerge from hidden areas is true. To fill these in a first step a digital terrain model (DTM), which represents the ground level without buildings or trees, is derived from the DSM. In the second step appropriate occluded regions are filled from this DTM.

Index Terms — digital surface models (DSM), DEM generation, urban DEMs, hierarchical area based matching, semi global matching, dense stereo algorithms, DEM analysis

2 DSM GENERATION

2.1 Algorithms used

The generation of digital surface models from stereo imagery is based on the correlation of many small regions of the two stereo images relative to each other. The relative distances are parallaxes which in turn give the absolute distance to the observer or height if the position and acquisition con-

ditions are known. Two main methods are investigated and analysed for Prism scenes in this paper:

- The so called classical hierarchical intensitiy based matching and
- Dense stereo methods based on dynamic line warping and semiglobal matching.

2.2 Hierarchical intensitiy based matching

The classical approach known as hierarchical intensity based matching consists of the following main steps [1]:

- Using the Foerstner interest operator [8] characteristic points are selected in one image of the stereo images pair
- These points are searched in the second stereo partner using an area based matching process
- The found matches are refined to subpixel accuracy by a local least squares adjustment

These steps are conducted in an image pyramid for coping with large distortions in the images.

Following the last step a modified region growing after [9] is applied to all found correlation points. The subpixel based region growing avoids proceeding into areas with low texture. After blunder elimination 3D points are derived by forward intersection using rigorous modeling of the imaging process or rational polynomial functions (RPC, see [10]). These 3D points are finally interpolated to a regular DSM.

2.3 Dense stereo by a fusion of Digital line warping and Semiglobal matching

Dense stereo algorithms were developed for computer vision purposes and are nowadays also applied to other stereo imagery. These algorithms are mostly based on epipolar images. In such images scan lines can be correlated directly line by line between the two images of the stereo image pair.

Therefore the satellite images has to be transformed to epipolar geometry following [11] in a first step. For the generation of the DSM the following algorithms were used:

- Semiglobal matching [5] and
- A fused algorithm consisting of dynamic line warping [6] and semiglobal matching.

In the dynamic line warping the two stereo images get warped onto each other line by line using a dynamic programming approach known from speech recognition:

- In a first step a cost matrix $M_{i,j}$ is built up from the two lines y of the stereo images I_1 and I_2:
  $M_{i,j} = | I_1(i,y) - I_1(j,y) |$
- Afterward $M_{i,j}$ is summed up to $D_{i,j}$ following the dynamic programming approach linearizing the recursive process of finding the path with the lowest cost

- Last a backtracking starting at the lowest rightmost element of $D - D_{w,w}$ with w as width of the stereo images – following the next lowest neighbour to the left, up or left-up a minimal path is built
- The deviations of this minimal path to the main diagonal of the matrix give the local distortions
- These distortions resemble parallaxes which in turn can be transformed to absolute height values by applying the acquisition geometry.

The semiglobal matching approach extends this algorithm by expanding the matrix M to a three dimensional matrix also containing all lines of the images in parallel. Afterwards the summarization of M to D is expanded to 8 or 16 two dimensional summarizations over the full 3D matrix M. In this case the backtracking of the minimal path doesn't work anymore and so the position of minimal values in disparity range give the searched parallaxes.

In the fused algorithm the summarization of only one scan line (dynamic line warping) or over all scan lines (semiglobal matching) is changed to a small region of ± 2 to 5 lines which are sufficient enough. Also a detection of occlusions is included.

3 DATA

The investigation is performed on an ALOS Prism scene covering Athens. The scene was acquired 2007-06-10 at 9:19 GMT with scene center at 38.021° north and 23.876° east.

Fig. 1: Nadir image of Athens scene, section 3750 m × 3750 m (1500 px × 1500 px) showing the Acropolis area (left from the image center)

The reference DSM was generated using the fused algorithm from an Ikonos stereo scene acquired 2004-07-24, 9:24 GMT with a ground resolution of 88 cm and viewing angles –19.99° and +13.17°.

Fig. 2: DSM of Ikonos Athens scene, section 500 m × 500 m showing the Acropolis area (left top)

4 RESULTS

The following figures show the results of the three matching algorithms in a section 1700 m × 1700 m around the Acropolis in UTM projection, Zone 34.

Fig. 3: DSM generated with the fused algorithm (digital line warping and semiglobal matching)

The DSM generated using the fused dense stereo method shows the occlusions resulting from steep walls which can not be seen in one of the stereo partners. Afterwards the digital terrain model (DTM) is generated from the DSM following [7] and the remaining occlusions get filled from the DTM.

For the filling of occlusions in future more sophisticated methods are recommended to distinguish between courtyards and streets (to be filled with ground values) and distortion effects on slanted surfaces (to be filled from neighboring values) as shown in Fig. 4.
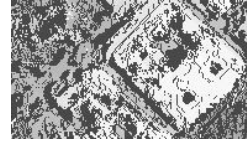

Fig. 4: Occlusions detected in small courtyards and on roofs (part of Ikonos reference DSM)

Fig. 5: DSM generated with the semiglobal matching method

Comparing the fused method DSM to the semiglobal matching DSM (see Fig. 5) some more fine details and less outliers can be detected. Also the detected occlusions allow a more sophisticated generation of true ortho images (not needed in the case of Prism, since here a nadir image exists).
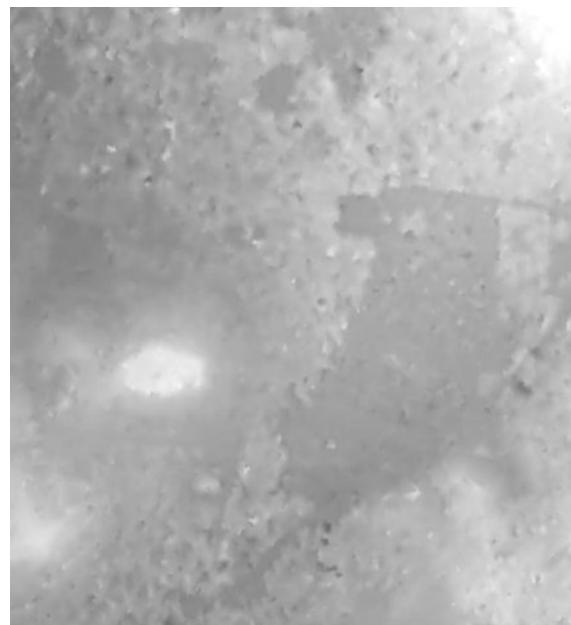

Fig. 6: DSM generated with the classical hierarchical area based matching method

The DSM generated with the classical hierarchical area based matching method also shows the struc-

ture of the urban area. But fine details like court-yards or narrow streets get interpolated since in occluded areas no matching can be performed and so these areas vanish.

The three resulting DSMs are compared in Fig. 8 for a section across the Acropolis (Fig. 7).
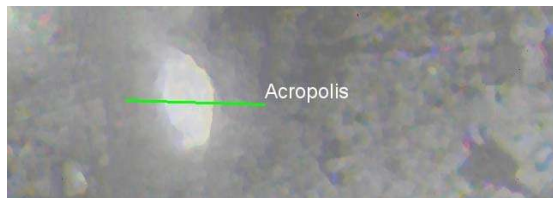

Fig. 7: Cross section across Acropolis

Fig. 8: Comparison of disparity values for classical method (blue), semiglobal matching (red) and fused method (green)

The large coincidence of the two dense stereo methods can be seen also showing some outliers in the case of the semiglobal matching. The classical method in contrast generates a relatively smooth DSM with an interpolated slant on the northern slope of the Acropolis since due to the occlusion situation no matching can be performed in this area.

Also the DSM generated with the fused method is compared to the reference DSM which results in the difference histogram shown in Fig. 9.
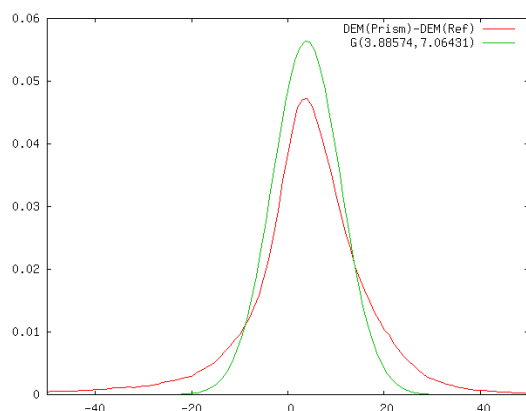
Fig. 9: Statistics of difference for DSM resulting from the fused method to reference DSM

The Prism DSM lies in average 3.9 meters above the reference DSM mostly due to the filling of small details. The standard deviation is 7 meters.

5 CONCLUSION

The ALOS Prism three line scanner stereo imagery is also a rather good base for the generation of relatively dense urban digital surface models. But for the generation of city models the usage of dense stereo algorithms rather than area based matching methods is strongly recommended since occlusions can not be detected in non dense stereo methods and so all occluded areas simply get interpolated. In contrast dense stereo methods – in the actual implementation the used fusion method – can detect occlusions and mark them for further post-processing.

6 REFERENCES

[1] Lehner, M. and Gill, R., 1992: Semi-automatic derivation of digital elevation models from stereo-scopic 3-line scanner data. ISPRS, 29 (B4), pp. 68–75.
[2] Collins T., 2004: Graph Cut Matching In Computer Vision, University of Edinburgh, February 2004.
[3] Pentenrieder, C., 2008. Analyse und Vergleich von 3D-Stereo-Verfahren für hochauflösende Satellitenbilder. Diploma theseis, Hochschule für angewandte Wissenschaften, München.
[4] Scharstein D., Szeliski R., 2002: A taxonomy and evaluation of dense two- frame stereo correspon-dence algorithms, International Journal of Com-puter Vision, 47(1/2/3):7-42, 2002.
[5] Hirschmüller H., 2005: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, IEEE Conference on Com-puter Vision and Pattern Recognition, San Diego USA, June 2005.
[6] Krauß, T., Reinartz, P., Lehner, M., Schroeder, M. and Stilla, U., 2005: DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. In: International Ar-chives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
[7] Krauß, T., Reinartz, P., U., 2009: Refinement of urban digital elevation models from very high resolution stereo satellite images. In: International Ar-chives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38 (1/W3). ISPRS Workshop, Hannover 2009.
[8] Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS Intercommission Workshop, Interlaken.
[9] Otto, G. and Chau, T., 1989: Region growing algorithm for matching of terrain images. Image and vision computing (7) 2, pp. 83–94.
[10] Jacobsen, K., Büyüksalih, G. and Topan, H., 2005: Geometric models for the orientation of high reso-lution optical satellite sensors. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
[11] Morgen M., 2004: Epipolar Resampling of Linear Array Scanner Scenes, PhD thesis, University of Calgary, UCGE Reports Nr. 20193, Canada, 2004.